View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Institutional Knowledge at Singapore Management University

Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection Lee Kong Chian School Of Business

Lee Kong Chian School of Business

7-2011

Size and Return: A New Perspective

Fangjian FU
Singapore Management University, fjfu@smu.edu.sg

Wei YANG
University of Rochester

Follow this and additional works at: https://ink.library.smu.edu.sg/lkcsb_research

Part of the Finance and Financial Management Commons, and the Portfolio and Security Analysis Commons

Citation

FU, Fangjian and YANG, Wei. Size and Return: A New Perspective. (2011). *China International Conference in Finance, Wuhan, July 2011*. 1-57. Research Collection Lee Kong Chian School Of Business.
Available at: https://ink.library.smu.edu.sg/lkcsb_research/3158

This Conference Paper is brought to you for free and open access by the Lee Kong Chian School of Business at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection Lee Kong Chian School Of Business by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email libIR@smu.edu.sg.

Size and Return: A New Perspective

Fangjian Fu[1] and Wei Yang[2]

[1]Lee Kong Chian School of Business, Singapore Management University, 50 Stamford Road, Singapore 178899. fjfu@smu.edu.sg.
[2]William E. Simon Graduate School of Business Administration, University of Rochester, Rochester, NY 14627. wei.yang@simon.rochester.edu.

We are grateful to Gurdip Bakshi, Jonathan Berk, John Campbell, John Cochrane, Darrell Duffie, Phil Dybvig, Wayne Ferson, Anisha Ghosh, Bruce Grundy, David Hirshleifer, Joachim Inkmann, Tim Johnson, John Long, Spencer Martin, Robert Novy-Marx, Lubos Pastor, Bill Schwert, Bill Sharpe, Robert Stambaugh, Sheridan Titman, Jerry Warner, Toni Whited, Guofu Zhou, Jerry Zimmerman, and seminar participants at University of Rochester, University of Connecticut, UIUC, the 21st Annual Conference on Financial Economics and Accounting, and the 2011 Finance Down Under Conference for helpful suggestions. We acknowledge the Best Paper Award at the 2011 Finance Down Under Conference.

Abstract

We document robust empirical evidence that, after controlling for idiosyncratic volatility, large stocks earn significantly higher returns than small stocks. Our empirical results indicate that idiosyncratic volatility is positively related to return, but negatively related to size. Hence, failure to control for idiosyncratic volatility generates a downward omitted variable bias and leads to the widely documented negative relation between size and return. We explain the two contrasting size-return relations, with and without the control for idiosyncratic volatility, in a parsimonious equilibrium model that incorporates three empirical regularities: some individual investors are under-diversified; small stocks have higher idiosyncratic volatilities than large stocks; and large stocks, relative to their size, are held by fewer investors than small stocks. Investors follow mean-variance optimization to allocate wealth among their stocks. To clear the markets, large stocks have to offer higher expected returns to induce their relatively smaller number of investors to allocate more of their wealth. This positive size-return relation is masked because small firms have higher idiosyncratic volatilities and therefore earn higher returns as a result of investor under-diversification.

Keywords: size effect; cross-section of stock returns; investor under-diversification

JEL classification codes: G12; G11

1 Introduction

Studies have documented that small stocks earn higher average returns than large stocks. This cross-sectional stock return pattern is often referred to as the "size effect".[1] In the sample of the stocks traded on the NYSE, AMEX, and NASDAQ during January 1926 to December 2009, we confirm this traditional negative size effect. Like Schwert (2003), we find that the size effect is stronger in early subsample periods and has largely disappeared since documented by Banz (1981) and Reinganum (1981).

More interestingly, we find that, after controlling for idiosyncratic volatility, the size effect becomes positive — large stocks earn significantly higher returns than small stocks. Specifically, in standard Fama-MacBeth regressions of stock returns, without idiosyncratic volatility as a regressor, the slope on size is negative; when idiosyncratic volatility is included as a control, the slope on size turns positive and highly significant. The positive size-return relation remains significant after controlling for other usual determinants of cross-sectional returns. The evidence is also robust in both early and later subsample periods. We further demonstrate the quantitative magnitudes of the positive size-return relation using portfolio sorts. In each month, we sort stocks first by idiosyncratic volatility and then by size into 10×10 portfolios, using the breaking points of NYSE stocks. Within the same idiosyncratic volatility deciles, the average return spreads between the largest and smallest size portfolios range between 0.64% and 2.60% per month. These large return spreads are not explained by the Fama-French three factors.

Our empirical results indicate that small firms have high idiosyncratic volatilities. We also confirm the findings in earlier studies that stocks with high idiosyncratic volatilities, or residual standard deviations, earn high returns. As a result, idiosyncratic volatility creates a negative link between size and return. Failure to control for idiosyncratic volatility results

[1] See Banz (1981) and Reinganum (1981) for the first finding of this effect, and Schwert (1983) for a survey of early studies. Fama and French (1992) and Schwert (2003) extend the analysis to more recent data.

in a downward omitted variable bias and leads to the widely documented negative relation between size and return.

We explain the two contrasting size-return relations, unconditional and conditional on idiosyncratic volatility, in a parsimonious equilibrium model. The model economy is populated with a large number of stocks of different size and an even larger number of investors. Most importantly, the model incorporates three empirical regularities: (1) some individual investors are under-diversified; (2) small stocks have high idiosyncratic volatilities; (3) large firms, relative to their size, are held by fewer investors than small firms. Evidence for the first empirical regularity is well documented.[2] For the second empirical regularity, the correlation between size and idiosyncratic volatility is -0.52 in our sample. We also provide evidence for the last empirical regularity in the paper. Specifically, we regress log number of shareholders on log market capitalization and the result suggests roughly a square-root dependence — a stock that is four times as large has on average about twice as many investors. In other words, while larger stocks are held by more investors, the relation between the number of investors and stock size is concave: the size-scaled number of investors decreases with stock size.

In the baseline setup, we focus on under-diversified individual investors. This is consistent with the observation that such investors play a dominant role in the stock market for a large part of our sample period. In recent decades, a growing number of individual investors hold stocks through institutions, and thus become diversified. To explore the implications of this development, we also investigate an extended setup which accommodates diversified individual investors by introducing large and diversified mutual funds into the model economy.

Specifically, in the baseline model, investors hold different portfolios of a small number of

[2] As surveyed in Campbell (2006), earlier studies find that the number of stocks held by a typical household or individual investor is only one or two. More recently, this number appears to increase to about four [Barber and Odean (2000) and Goetzmann and Kumar (2008)].

stocks, and allocate wealth among their stocks following mean-variance optimization. The expected return of a stock is determined in the equilibrium so that the aggregated demand from investors holding the stock equals the supply, i.e., the market capitalization of the stock. We calibrate the model parameters so that the model reproduces the salient quantitative features of the empirical data. In our baseline calibration, the model economy contains 2,000 stocks and 200,000 investors each holding 4 stocks. We solve the model numerically and obtain the cross sections of size, idiosyncratic volatility, and expected return in simulated model economies.

The model generates the same patterns as documented in the empirical data. High idiosyncratic volatility stocks earn higher returns than low idiosyncratic volatility stocks. Without the control for idiosyncratic volatility, small firms exhibit higher returns than large firms, i.e., the traditional size effect. After controlling for idiosyncratic volatility, the size-return relation turns positive.

To identify the economic mechanisms underlying the relations between size, idiosyncratic volatility, and return, we conduct counterfactual experiments on the model. Specifically, we make changes, one at a time, on each of the three empirical regularities incorporated in the model. If we set the size-scaled number of investors to be the same for large and small firms (i.e., the number of investors increases linearly with stock size), the size-return relation conditional on idiosyncratic volatility becomes flat. If we increase the number of stocks held by investors (i.e., investor portfolios are more diversified), both the relation between idiosyncratic volatility and return and the relation between size and return controlling for idiosyncratic volatility become less positive. If we set zero correlation between size and idiosyncratic volatility, the model yields a strong positive relation between size and return, with or without the control for idiosyncratic volatility.

These experiments suggest that the positive size-return relation results from the joint effect of investor under-diversification and the decreasing size-scaled number of investors.

Although large stocks are held by more investors, they have, when scaled by size, fewer investors than small stocks. Consequently, in equilibrium, large stocks have to offer higher expected returns to induce their relatively smaller number of investors to allocate more of their wealth. These experiments also confirm that the positive relation between idiosyncratic volatility and return is driven by investor under-diversification. Idiosyncratic volatility contributes to the risk of under-diversified portfolios, and thus under-diversified investors demand higher returns for stocks with higher idiosyncratic volatilities. Lastly, because small firms have higher idiosyncratic volatilities and thus earn higher returns than large firms, this gives rise to a negative link between size and return via idiosyncratic volatility. The positive size-return relation due to the decreasing size-scaled number of investors is masked by this negative link. Controlling for idiosyncratic volatility unveils this mask.

As robustness checks, we investigate a few variants of the baseline model. We impose restrictions on investor shorting and borrowing, and relax the assumption that all investors hold equal wealth. We increase the number of stocks as well as the number of investors. The results from these variants are qualitatively the same and quantitatively similar to those of the baseline model.

Finally, we investigate an extension of the baseline model, in which the economy contains under-diversified individual investors as well as large and diversified mutual funds. This extended setup reflects the growing trend over the recent decades among individual investors to hold stocks through institutions and thus become diversified. The results from the extended model remain qualitatively the same, though smaller in magnitude. This, of course, is anticipated since the relations between size, idiosyncratic volatility, and return critically depend on investor under-diversification.

Our model is a return to the tradition of the classic studies of Sharpe (1964) and Lintner (1965) on the CAPM, and the seminal papers of Levy (1978) and Merton (1987) on the impact of investor under-diversification. In our model, like in these classic studies, investors allocate wealth among their stocks following mean-variance optimization, and stocks are economic commodities whose prices are determined by crossing supply and demand in the equilibrium. This equilibrium approach reveals important insights that do not readily transpire in the factor pricing framework.

Based on under-diversified investors' demand equations for stocks, Merton (1987) also suggests a positive size-return relation. However, the positive relation in our model is different from his. In Merton (1987), the relation is obtained from the partial derivative of investors' demand equation with respect to size, holding everything else fixed. In our model, the positive size-return relation arises in the equilibrium cross section of expected returns as a result of the decreasing size-scaled number of investors. To clear the markets, large stocks offer higher expected returns to induce more demand from their relatively smaller number of investors. This intuition has not yet been proposed in the existing literature. Moreover, our model incorporates key empirical regularities and replicates the cross-sectional joint distributions of size, idiosyncratic volatility, and the number of investors in the empirical data. The solutions of the simulated model economies demonstrate the equilibrium cross section relations between size, idiosyncratic volatility, and stock return, which can be directly compared with the real data.

In addition, the focus of Merton (1987) is not on the relation between size and return. Instead, he illustrates the importance of investor recognition (or investor base) on stock returns. Inspired by his work, a number of empirical studies provide supporting evidence to his prediction that broader investor recognition is associated with lower expected returns.[3] The investor recognition literature often uses the number of shareholders to measure how well-known a stock is. Our study uses the number of shareholders *scaled by size* to reflect investor demand for stocks. These two quantities, while related, are meant to capture very

[3]See, among others, Kadlec and McConnell (1994), Amihud, Mendelson, and Uno (1999), Foerster and Karolyi (1999), Gervais, Kaniel, and Mingelgrin (2001), and Dyl and Elliott (2006), and Bodnaruk and Ostberg (2009).

different economic concepts. Large stocks are known to and held by more investors than small stocks. The increase in the number of investors with size, however, is not linear, but concave. The investor recognition literature and our paper address distinct research questions. Their focus is on the relation between investor recognition and stock returns, while we examine how the concavity plays an important role in explaining the size-return relation in the cross section.

The economic mechanism proposed in our paper for the size-return relation depends in part on the positive, contemporaneous relation between idiosyncratic volatility and return. This positive relation is documented empirically in Friend, Westerfield, and Granito (1978) and Levy (1978), which we confirm in an extended sample. The theoretical intuition that investor under-diversification drives the positive relation is demonstrated in Levy (1978) and Merton (1987), which we also verify in our model.[4] In one of the counterfactual experiments, when we increase the number of stocks held by investors while keeping everything else the same, the positive relation between idiosyncratic volatility and return becomes weaker. This vividly demonstrates how investor diversification affects the pricing of idiosyncratic risk. In addition, as mentioned above, the relation between idiosyncratic volatility and return becomes less positive when mutual funds are introduced into the model.

Our paper also suggests a potential explanation for the widely documented negative relation between size and return. Namely, it results from an omitted variable bias due to the failure to control for idiosyncratic volatility. This explanation provides an interesting

[4] In contrast to the positive, contemporaneous relation, Ang, Hodrick, Xing, and Zhang (2006) report a negative relation between return and the previous-month idiosyncratic volatility. Fu (2009) suggests that the contrasting results are due to the fact that idiosyncratic volatility varies substantially from month to month, and thus the idiosyncratic volatility of the previous month could be very different from the idiosyncratic volatility for the current month. Moreover, Fu (2009) shows that, consistent with the positive relation, high idiosyncratic volatilities in the previous month are accompanied with high returns in the same month. The low returns in the following month observed by Ang, Hodrick, Xing, and Zhang (2006) are largely due to the subsequent month return reversal documented in Jegadeesh (1990). We therefore follow the classic studies to use the contemporaneous idiosyncratic volatility that reveals the positive relation between return and idiosyncratic volatility.

alternative to the economic insights on the size effect highlighted in a number of influential studies such as Chan and Chen (1991), Berk (1995), Berk, Green, and Naik (1999), and Gomes, Kogan, and Zhang (2003).

In the rest of the paper, we first discuss the data in Section 2. Section 3 reports the empirical evidence on stock returns and the decreasing size-scaled number of investors. We present the model in Section 4 and discuss the calibration of parameters. In Section 4.5, we present the numerical solution of the model equilibrium, discuss the results, and explore the underlying economic intuition. The concluding section summarizes the key insights and proposes potential extensions of our study.

2 Data and variables

Our full sample consists of the stocks traded on the NYSE, AMEX, and NASDAQ during the period of January 1926 to December 2009, with 3,549,169 firm-month observations in total. Stock returns (RET) are monthly, and XRET is the return in excess of the one-month T-bill rate. Market capitalization (ME) is the product of the end-of-month closing price and the number of shares outstanding, adjusted by the Consumer Price Index and expressed in millions of year 2000 dollars. Idiosyncratic volatility (IVOL) is the standard deviation of the residuals from regressing individual stock returns on market returns. Specifically, for stock i, in month t, we run a time-series regression of the daily stock returns on the contemporaneous and three lagged value-weighted market returns:

$$\begin{aligned}\mathrm{RET}_{i,\tau} = \alpha_{i,t} + \beta_{0,i,t}\mathrm{MRET}_{\tau} + \beta_{1,i,t}\mathrm{MRET}_{\tau-1} \\ + \beta_{2,i,t}\mathrm{MRET}_{\tau-2} + \beta_{3,i,t}\mathrm{MRET}_{\tau-3} + \varepsilon_{i,\tau}, \quad \text{day } \tau \in \text{ month } t.\end{aligned} \tag{1}$$

Here, $\mathrm{RET}_{i,\tau}$ is the return of stock i on day τ in month t, and MRET is the market return. We compute IVOL by multiplying the standard deviation of the regression residuals with the square root of the number of trading days in month t. The use of the lagged market returns is to adjust for the effect of non-synchronous trading [Dimson (1979)].

As specified above, idiosyncratic volatility is estimated using the CAPM. This is close in spirit to the theoretical model presented later, in which the economy is driven by a single macroeconomic factor. In unreported robustness checks, we also find similar empirical results if idiosyncratic volatility is measured using the Fama-French three-factor model.

Panel A of Table 1 reports the descriptive statistics of the sample. Since ME and IVOL are positively skewed, we take natural logarithm and report the summary statistics of log(ME) and log(IVOL) as well. Table 2 reports the time-series averages of the cross-sectional simple correlations between the stock return, log(ME), and log(IVOL). The correlation between size and return is negative and significant. This is consistent with the traditional size effect — small firms earn higher returns than large firms. The correlation between return and IVOL is significantly positive; high idiosyncratic volatility is associated with high return. Finally, the correlation between size and IVOL is -0.52, negative and statistically significant: small firms have more volatile stock returns.

Panel B of Table 1 summarizes CSHR, the number of common shareholders. The fundamental annual file of Compustat reports this data item since fiscal year 1975. The statistics for this variable is based on 201,062 firm-year observations during the period 1975–2009. The variable CSHR is positively skewed. We also take log transformation and report the summary statistics of log(CSHR).

3 Empirical evidence

We first investigate the relations between firm size, idiosyncratic volatility, and return, and then the relation between the number of shareholders and firm size.

3.1 Fama-MacBeth regressions of stock returns

We use the Fama-MacBeth regressions to examine the empirical relations between size, idiosyncratic volatility, and return. Each month, we regress individual stock excess returns on lagged log(ME), on log(IVOL), and on both variables, respectively. We report the time-series averages of the slope coefficients and compute the t-statistics as the time-series averages of the slopes divided by the corresponding time-series standard errors. We examine the full sample period 1926:01–2009:12 and two subperiods: the early period 1926:01–1967:12, and the later period 1968:01–2009:12. The results are reported in Table 3.

For the unconditional relation between size and return, we regress excess stock returns on lagged log(ME) and obtain negative and significant slopes. Hence, consistent with existing studies, small firms exhibit higher returns than large firms. Existing studies also suggest that the relation between size and return varies over time.[5] The results in Table 3 indicate that the slope on size is smaller in magnitude in the later sample period

For the unconditional relation between idiosyncratic volatility and return, regressing excess stock returns on log(IVOL) obtains positive and significant slopes. The results confirm the findings of Friend, Westerfield, and Granito (1978) and Levy (1978). The slopes are significant in both subperiods, while the magnitude decreases over time.

The key interest of our study is the size-return relation after controlling for idiosyncratic volatility. As shown in Table 3, when both log(ME) and log(IVOL) are included as regressors, the slope on log(ME) becomes positive and significant. That is, controlling for idiosyncratic

[5]See the survey in Schwert (2003).

volatility, stock returns are positively related to size. The results for the two subsample periods also suggest that the slope decreases over time.

Studies have shown that a number of variables are related with cross-sectional stock returns. To investigate the robustness of the empirical results on the relations between size, idiosyncratic volatility, and return, we include additional variables in the Fama-MacBeth regressions. These additional variables include the ratio of book-to-market equity, liquidity and its variance,[6] and past returns. We follow Fama and French (1992) for the ratio of book-to-market equity, Chordia, Subrahmanyam, and Anshuman (2001) for liquidity and its variance, Jegadeesh (1990) for the preceding month returns, and Jegadeesh and Titman (2001) for the past 6-month (skipping the preceding month) returns. Some of these additional variables are not available prior to 1960s, and thus we report the regression results for the later subperiod of 1968:01–2009:12. As shown in Table 4, with these additional controls, the key empirical results remain qualitatively the same and quantitatively similar: without the control for idiosyncratic volatility, the size-return relation is negative; after controlling for idiosyncratic volatility, the size-return relation becomes positive.

The dependent variables in the regressions above are stock excess returns XRET. As indicated in Table 1, XRET is somewhat right-skewed, with a skewness of 1.69 in the full sample. In contrast, the skewness of log excess returns, log(100+XRET), is -0.56 in the full sample, and -0.60 in the later subperiod of 1968:01–2009:12. As a robustness check, Table 4 also reports the regression of log excess returns on log(ME), log(IVOL), and the additional control variables. The results again confirm our key result: controlling for idiosyncratic volatility, the relation between size and return is positive and significant.

[6] Because large stocks tend to be more liquid, the implication of the liquidity effect — less liquid stocks earn higher returns — is opposite to our finding of the positive size-return relation. This intuition suggests that our finding is not due to the liquidity effect, which is confirmed by the empirical evidence.

3.2 Portfolio sorting

We use portfolio sorting to demonstrate the quantitative magnitudes of the empirical relations documented in the Fama-MacBeth regressions.

The results on size portfolios are presented in Table 5. In each month, we sort stocks into deciles based on ME of the previous month, and following Fama and French (1992), we use the breaking points based on the ME of NYSE stocks only. We compute the equal- and value-weighted portfolio returns and report the time-series averages. Besides the full sample period and the early and later subperiods, we also examine the most recent period of 1982:01–2009:12, which is after the documentation of the size effect in Banz (1981) and Reinganum (1981). Consistent with existing studies and confirming the regression results presented earlier, small stocks earn higher average returns. In addition, the return spread between the largest and the smallest ME deciles is much more negative in the early subsample period than in the later period, and becomes insignificant during the most recent decades.[7]

Table 6 presents the results on idiosyncratic volatility sorted portfolios. In each month, we sort stocks into deciles by IVOL, using the breaking points based on the IVOL of NYSE stocks only. Confirming the regression results, stocks with high idiosyncratic volatilities earn higher returns than stocks with low idiosyncratic volatilities. Further, comparison across the two subperiods indicates that the return spread decreases over time.

To demonstrate the size-return relation conditional on idiosyncratic volatility, we employ a sequential sorting procedure to form stock portfolios that have similar idiosyncratic volatility but very different size. In each month, we first sort stocks into deciles based on IVOL, and then sort the stocks in each IVOL decile into 10 portfolios based on ME of the previous month. In both steps of sorting, we use the breaking points based on NYSE stocks only. The purpose of the first sort is to narrow down the variation of IVOL, while the second

[7]Schwert (2003) suggests the disappearance of the negative size-return relation could be due to increased arbitrage activities since the finding of the size effect.

sort separates the stocks with similar IVOL by size.

For each month, this sequential sorting yields 100 portfolios. Table 7 presents the time-series averages of median ME and median IVOL for the 100 portfolios. Panel A shows that within each IVOL decile, stocks in different ME portfolios exhibit very different size. Panel B demonstrates that the sequential sorting effectively controls for idiosyncratic volatility across size portfolios. For IVOL deciles 1 to 9, the spreads of median IVOL between the largest and the smallest size portfolios are all below 1%. In other words, within IVOL deciles 1 to 9, there is little variation in IVOL across the size portfolios.

The highest IVOL decile is an exception. Panel B shows that the median IVOL decreases by about 10% from size deciles 1 to 10. In other words, for the size portfolios in the highest IVOL decile, the sequential sorting does not achieve an effective control for idiosyncratic volatility. Hence, the interpretation of the portfolio return results for this IVOL decile requires special attention.

We compute the equal- and value-weighted excess returns in each month for the 100 portfolios and Table 8 reports the time-series averages for the full sample period. Within each of IVOL deciles 1 to 9, the average portfolio returns increase with size; the return spreads between size deciles 10 and 1 are positive and statistically significant. For the full sample, the average monthly return spreads range between 0.64% and 2.60% for equal-weighted portfolio returns. In addition, we run the time-series regressions of the return spreads on the Fama-French three factors. As reported in the last column of Table 8, the estimated intercepts, or the alphas, are positive and statistically significant. As noted earlier, for IVOL deciles 1 to 9, the control for IVOL across the size portfolios is effective. Hence, the positive return spreads suggest a positive relation between size and return.

In the highest IVOL decile, however, the return decreases with size, and the return spread between ME deciles 10 and 1 is negative and significant. As noted earlier, there is a substantial decrease in IVOL across the size portfolios in the highest IVOL decile. Since idiosyncratic volatility and return are positively correlated, if the effect of the decreasing IVOL dominates, this can give rise to the negative return spread in the highest IVOL decile.

In addition, although the highest IVOL decile consists of about 26% of the stocks in number, it supplies less than 3% of the total market capitalization.[8] The economic importance of this decile is likely small.[9]

For the subsample periods of 1926:01–1967:12 and 1968:01–2009:12, Panel B of Table 8 reports the return spreads between the largest and smallest size portfolios for each of the 10 IVOL deciles. The results are similar to those for the full sample period: for IVOL deciles 1 to 9, the return spreads are positive and significant; for IVOL decile 10, the return spread is negative. The results also indicate smaller magnitudes for the return spreads in the later period.

3.3 Omitted variable bias

In summary, our empirical results confirm the widely documented negative relation between size and return, and further, demonstrate that after controlling for idiosyncratic volatility, the relation becomes positive. These empirical results suggest a potential explanation of the conventional negative size effect. That is, failure to control for idiosyncratic volatility results in an omitted variable bias, which leads to the negative relation. The bias is downward

[8] The numbers of stocks are different across the IVOL deciles because the stocks are sorted using the breaking points based on NYSE stocks only. Consistent with the negative correlation between size and idiosyncratic volatility, the highest IVOL decile contains very small stocks.

[9] We conduct additional robustness checks. The highest IVOL decile contains very small stocks. Small stocks with high return volatilities have a higher probability to be delisted in the following period than large stocks. The CRSP's monthly stock return file does not include delisting returns. This creates a survivorship bias, which has been shown contributing to the negative return spread between large and small stocks [Shumway and Warther (1999)]. When we include delisting returns (obtained from CRSP's monthly stock event file) in computing the portfolio returns, the return spread between the largest and smallest size portfolios in the highest IVOL decile indeed becomes less negative. Including delisting returns has little impact on the return spreads between the size portfolios in the other IVOL deciles.

We also find that after excluding January stock returns, the portfolio return results are similar to those in Panel A of Table 8. As a matter of fact, the return spreads between the largest and the smallest size portfolios in IVOL deciles 1 to 9 become more positive, while the negative spread in IVOL decile 10 becomes smaller in magnitude.

because idiosyncratic volatility is negatively related to size, but positively related to return, thus creating a negative link between size and return.

This downward bias is explained in more detail in the appendix. Specifically, two channels contribute to the unconditional relation between size and return. The first channel is the positive size-return relation conditional on idiosyncratic volatility, which we document in this study. In the second channel, size is negatively correlated with idiosyncratic volatility, which, in turn, is positively associated with return. This negative, second channel gives rise to the downward bias, and it more than offsets the positive, first channel, resulting in the unconditional negative relation between size and return. In other words, the positive relation is masked by the negative link between size and return via idiosyncratic volatility.[10] Controlling for idiosyncratic volatility removes the mask.

3.4 Number of individual investors

In this section, we illustrate the empirical relation between the number of investors and stock size. Our measure for the number of investors is CSHR, the number of common shareholders reported in the fundamental annual file of Compustat since fiscal year 1975. As perhaps the only source of information on the number of shareholders for a large panel of US firms, CSHR is frequently used in the investor recognition literature to measure how well-known a stock is. Our study, however, focuses on its relation with stock size.

For each year of 1975–2009, we run a cross-sectional regression of log(CSHR) on log(ME). Table 9 reports that the time-series average of the slope on log(ME) is 0.42, and the time-series average of R^2 is about 34%. A positive slope suggests that larger firms are held by more investors. This is consistent with the intuition that larger firms tend to be more well-known, and thus attract more investors. If the slope is 1, it implies that the number of

[10]Similarly, for the unconditional relation between idiosyncratic volatility and return, there is also a downward bias because size is the omitted variable. The bias weakens the positive relation, but is not strong enough to flip the sign. See the appendix for more details and Table 3 for the empirical results.

investors increases linearly with firm size, or equivalently, the size-scaled number of investors is the same for all firms. A slope between 0 and 1 suggests that, while larger stocks are held by more investors, the relation between the number of investors and stock size is concave: the size-scaled number of investors decreases with stock size.

Due to the inclusion of institutional investors, CSHR as the proxy for the number of individual investors involves measurement errors. Evidence from the 13f institutional holding dataset, however, suggests that the resulting impact is likely small. The increasing predominance of institutional investors is a phenomenon less than three decades old. In 1980, the total number of institutional investors is only about 500; the median institutional ownership is below 10% for NYSE stocks and zero for NASDAQ stocks. In other words, the influence of institutional investors is rather small in the early years of our sample. Nonetheless, we still find that the slope coefficients are about 0.50 in the late 1970s to early 1980s. In addition, we adjust both CSHR and ME to exclude the effect of institutional investors — we subtract the number of institutional investors from CSHR, and the market capitalization held by these institutional investors from ME — and then run regressions of log adjusted CSHR on log adjusted ME. The time series average of the slope coefficients increases only slightly to 0.46, and the average R^2 is about the same.

To summarize, the empirical results indicate that, relative to their size, large firms are held by fewer investors, or equivalently, the size-scaled number of investors decreases with firm size. As shown subsequently in the model, this empirical fact together with investor under-diversification generate a positive relation between size and return. The smaller the slope coefficient calibrated for the relation between log(CSHR) and log(ME) (i.e., the stronger the concavity), the larger the magnitude of the model results. Our model calibration will use a conservative value of 0.5, which implies a square-root relation between the number of investors and firm size: a firm four times as large has on average about twice as many investors.[11]

4 Model

In this section, we present a parsimonious equilibrium model of many stocks and numerous investors. Our model builds on those in the classical CAPM literature and the seminal studies of Levy (1978) and Merton (1987). In the model economy, investors are mean-variance optimizers over their stocks, and the expected return of each stock is determined so that the aggregated demand equals the supply. Further, we incorporate into the model three empirical regularities observed in the data — some individual investors are under-diversified, large firms have lower idiosyncratic volatilities than small firms, and the size-scaled number of investors decreases with firm size.

In the baseline setup, we focus on under-diversified individual investors. This is consistent with the observation that such investors play a dominant role in the stock market for a large part of our sample period. In recent decades, a growing number of individual investors hold stocks through institutions, and thus become diversified. To explore the implications of this development, we investigate an extended setup which accommodates diversified individual investors by introducing large and diversified mutual funds into the model economy.

We calibrate the model to match the salient quantitative properties of the empirical data, and explore the model implications on the cross section of stock returns. As shown subsequently, with investor under-diversification, the model generates a positive relation between idiosyncratic volatility and return. More importantly, the decreasing size-scaled number of investors and investor under-diversification generate a positive relation between size and return. These results, together with the negative correlation between size and idiosyncratic

[11]Here is an example to illustrate the decreasing size-scaled number of investors. At the end of fiscal year 1975, Eastman Kodak has a market capitalization of \$17.12 billion and is owned by 237.5 thousand investors; Xerox has a market capitalization of \$4.05 billion and is owned by 135.6 thousand investors.

volatility, can explain the empirical relations between size, idiosyncratic volatility, and return as documented in the paper.

4.1 Stocks

The model has two periods, time 0 and time 1, and consistent with our empirical analysis, the interval is one month. The economy is driven by a single macroeconomic factor

$$\tilde{F}_i = \sigma_F \tilde{f}, \qquad \tilde{f} \sim N(0,1). \tag{2}$$

The factor has a mean of 0 and a standard deviation of σ_F, and the factor shock $\tilde{f}$ is a standard normal random variable.

The economy is populated with a large number of stocks. Stock i pays out a random cash flow at time 1,

$$\tilde{C}_i = C_i(1 + B_i \tilde{F} + \sigma_i \tilde{\varepsilon}_i), \qquad \tilde{\varepsilon}_i \sim N(0,1), \qquad \mathrm{corr}[\,\tilde{\varepsilon}_i, \tilde{f}\,] = 0. \tag{3}$$

Here, C_i is the mean, B_i is the exposure of the cash flow to the macroeconomic factor, and σ_i is the standard deviation of the stock specific shock $\tilde{\varepsilon}_i$, which is a standard normal random variable. The factor shock $\tilde{f}$ and the stock specific shock $\tilde{\varepsilon}_i$ are independent.

Stock i is thus characterized by three parameters, C_i, B_i, and σ_i. The macroeconomic factor will ultimately give rise to the systematic risk in the economy, and B_i largely determines the loading of stock i on this risk. The stock specific shocks generate idiosyncratic risks, and σ_i largely determines the magnitude of idiosyncratic volatility of stock i. Finally, because of the short horizon between the two periods, the magnitude of the cash flow C_i largely determines the market capitalization of stock i.

We assume that cross-sectionally (across i), $\log C_i$, B_i, and $\log \sigma_i$ follow normal distributions. In addition, $\log C_i$ and $\log \sigma_i$ are correlated:

$$\text{corr}[\,\log C_i, \log \sigma_i\,] = \rho. \tag{4}$$

A negative ρ will generate a negative correlation between size and idiosyncratic volatility in the model, consistent with the empirical evidence. The distribution of B_i is independent of $\log C_i$ and $\log \sigma_i$. As a result, in the model the CAPM beta varies essentially independently from size or idiosyncratic volatility, and thus cannot explain the return patterns associated with size or idiosyncratic volatility. The stock specific shocks are correlated, and $\text{corr}[\,\tilde{\varepsilon}_{i1}, \tilde{\varepsilon}_{i2}\,]$ is drawn from a normal distribution.

Let V_i denote the time-0 present value of the time-1 random cash flow $\tilde{C}_i$ of stock i; V_i is also firm size or market capitalization. The gross return is

$$\tilde{R}_i = \frac{C_i}{V_i}(1 + B_i \sigma_F \tilde{f} + \sigma_i \tilde{\varepsilon}_i). \tag{5}$$

The expected gross return is simply

$$R_i = E[\,\tilde{R}_i\,] = \frac{C_i}{V_i}. \tag{6}$$

It then follows that

$$\tilde{R}_i = R_i(1 + B_i \sigma_F \tilde{f} + \sigma_i \tilde{\varepsilon}_i). \tag{7}$$

The covariance between two stocks i_1 and i_2 is

$$\text{cov}[\,\tilde{R}_{i1}, \tilde{R}_{i2}\,] = R_{i1} R_{i2} \big(B_{i1} B_{i2} \sigma_F^2 + \sigma_{i1} \sigma_{i2} \,\text{cov}[\,\tilde{\varepsilon}_{i1}, \tilde{\varepsilon}_{i2}\,]\big). \tag{8}$$

4.2 Under-diversified, mean-variance optimizing investors

The economy is populated with a large number of individual investors. A large literature documents that individual investors hold under-diversified portfolios,[12] and proposes numerous explanations.[13] To sharpen the focus of our study, we abstract from specific mechanisms underlying investor under-diversification, and take this empirical regularity as given. In the baseline model, similar to Levy (1978) and Merton (1987), we assign a small number of random stocks to each investor. Moreover, as detailed below, we assign stocks so that the model economy replicates the empirical relation between the number of investors and firm size.

In the baseline model, investors have the same wealth, and each of them makes a small number, M, of picks from the entire universe of stocks with replacement. The probability of stock i being picked is

$$P_i \propto C_i^{\lambda} e^{\sigma_\pi \pi_i}, \qquad \pi_i \sim N(0,1), \quad \lambda > 0. \tag{9}$$

The probability is proportional to the magnitude of the cash flow raised to the power of λ, and is also subject to a log-normal variation with σ_π as the standard deviation. The log-normal term represents factors that drive investors' picks but are orthogonal to C_i.

The same stock can be picked more than once by an investor. With a large number of stocks, duplication only occurs for stocks with very large P_i and in the portfolios of a tiny fraction of investors. Most investors' portfolios contain M different stocks.[14]

[12] See, among others, Blume and Friend (1975), Kelly (1995), Barber and Odean (2000), Polkovnichenko (2005), Calvet, Campbell, and Sodini (2007), Goetzmann and Kumar (2008), and the survey in Campbell (2006).

[13] See, among others, Brennan (1975), Kraus and Litzenberger (1976), Bloomfield, Leftwich, and Long (1977), Merton (1987), Odean (1999), Harvey and Siddique (2000), Shefrin and Statman (2000), Polkovnichenko (2005), Barberis and Huang (2008), Cohen (2009), Liu (2009), and Van Nieuwerburgh and Veldkamp (2010).

[14] As shown later, our calibration specifies a large number of investors, and as a result, all stocks are picked by some investors.

The total number of investors holding stock i is proportional to the probability P_i:

$$N_i \approx \text{const} \times C_i^{\lambda} e^{\sigma_\pi \pi_i}, \tag{10}$$

or

$$\log N_i \approx \text{const} + \lambda \log C_i + \sigma_\pi \pi_i. \tag{11}$$

Because the short horizon between the two periods in the model, the distribution of C_i largely determines that of V_i. Hence, this stock picking scheme allows the model to replicate the empirical relation reported earlier between CSHR and ME.

After picking stocks, each investor solves a mean-variance portfolio problem. For expositional clarity, we will suppress the investor index in the following. Let $\boldsymbol{R}$ denote the vector of the expected gross returns of the stocks in an investor's portfolio, and $\boldsymbol{\Sigma}$ be the covariance matrix of the stock returns. The investor can also borrow or lend a risk-free asset, with the gross risk-free rate R_f. Let $\boldsymbol{\omega}$ denote the vector of the weights for the stocks. The investor maximizes

$$(1 - \boldsymbol{\omega}'\mathbf{1})R_f + \boldsymbol{\omega}'\boldsymbol{R} - \frac{\delta}{2}\boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega}. \tag{12}$$

Here, δ is a preference parameter that determines the investor's mean-variance tradeoff. Without any constraints, the first order condition with respect to $\boldsymbol{\omega}$

$$\boldsymbol{R} - R_f\mathbf{1} - \delta\boldsymbol{\Sigma}\boldsymbol{\omega} = 0 \tag{13}$$

yields

$$\boldsymbol{\omega} = \frac{1}{\delta}\boldsymbol{\Sigma}^{-1}(\boldsymbol{R} - R_f\mathbf{1}). \tag{14}$$

If the investor is subject to constraints on shorting

$$\boldsymbol{\omega} > 0, \tag{15}$$

or borrowing up to, e.g., 30% of the net worth of the portfolio,

$$\boldsymbol{\omega}'\mathbf{1} \leq 1.3, \tag{16}$$

then the optimization is generally a quadratic programming problem.

4.3 Equilibrium

We set up the economy, the stocks, and the investors' holdings, and then solve for the equilibrium. We begin with an initial guess of expected stock returns. Using expected stock returns, we compute market capitalization V_i. The sum of all V_i is the total wealth in the economy, which is equally divided among the investors. Then we solve the portfolio problem for each investor. For each stock, we sum up the wealth invested in the stock by all the investors holding the stock. The total wealth invested in stock i, W_i, represents the demand, while the supply is V_i. Hence, if $W_i < V_i$, or the demand is less than the supply, we increase the expected return R_i, which induces investors holding stock i to allocate more of their wealth to the stock. Conversely, if $W_i > V_i$, or the demand is more than the supply, we decrease R_i, and as a result, investors holding stock i allocate less of their wealth to the stock. In equilibrium, the demand equals the supply for all the stocks. When the markets are clear for all the stocks, by the Walras law, borrowing and lending between investors at the risk-free rate also sum to 0.

With the solution of the equilibrium, the aggregate stock market return is

$$
\begin{aligned}
\tilde{R}_M = \frac{\sum_i V_i \tilde{R}_i}{\sum_i V_i} &= \frac{\sum_i V_i R_i (1 + B_i \sigma_F \tilde{f} + \sigma_i \tilde{\varepsilon}_i)}{\sum_i V_i} && (17)\\
&= \frac{\sum_i C_i}{\sum_i V_i} + \frac{\sum_i C_i B_i}{\sum_i V_i} \sigma_F \tilde{f} + \frac{\sum_i C_i \sigma_i \tilde{\varepsilon}_i}{\sum_i V_i} && (18)\\
&= \frac{\sum_i C_i}{\sum_i V_i} \left(1 + \frac{\sum_i C_i B_i}{\sum_i C_i} \sigma_F \tilde{f} + \frac{\sum_i C_i \sigma_i \tilde{\varepsilon}_i}{\sum_i C_i} \right) && (19)\\
&= R_M (1 + B_M \sigma_F \tilde{f} + \tilde{\varepsilon}_M), && (20)
\end{aligned}
$$

where the expected market return is

$$
R_M = \frac{\sum_i C_i}{\sum_i V_i}. \tag{21}
$$

The market return variance is

$$
\operatorname{var}[\tilde{R}_M] = R_M^2 \Big(B_M^2 \sigma_F^2 + \operatorname{var}[\, \tilde{\varepsilon}_M \,] \Big). \tag{22}
$$

The covariance between the market return and the return of stock i is

$$
\operatorname{cov}[\, \tilde{R}_i, \tilde{R}_M \,] = R_i R_M \Big(B_i B_M \sigma_F^2 + \sigma_i \operatorname{cov}[\, \tilde{\varepsilon}_i, \tilde{\varepsilon}_M \,] \Big). \tag{23}
$$

We can then compute the CAPM beta

$$
\beta_i = \frac{\operatorname{cov}[\, \tilde{R}_i, \tilde{R}_M \,]}{\operatorname{var}[\, \tilde{R}_M \,]}, \tag{24}
$$

and the idiosyncratic volatility

$$H_i = \sqrt{\text{var}[\, R_i \,] - \beta_i^2 \, \text{var}[\, \tilde{R}_M \,]}. \tag{25}$$

For the model, all variables are explicitly computed from the solution, while their empirical counterparts are estimated from the real data.

4.4 Model parameters

We choose the model parameters so that the model economy replicates the salient properties of the empirical data. In the baseline calibration, we set the number of stocks at 2000,[15] and the number of investors at 2×10^5. As surveyed in Campbell (2006), earlier studies find that the number of stocks held by a typical household or individual investor is only one or two. More recently, this number appears to increase to about four [Barber and Odean (2000) and Goetzmann and Kumar (2008)].[16] In our baseline calibration, each investor makes four stock picks, and we find that on average, 99% of investors end up holding four different stocks.

For the distribution of $\log C_i$, the mean is 5.01 and the standard deviation is 1.75. These two parameters are chosen so that in the model the mean and the standard deviation of $\log V_i$ match those of log(ME) of the firms in the empirical data. For the distribution of $\log \sigma_i$, the mean is 2.14 and the standard deviation is 0.65. These two parameters are chosen so that in the model the mean and the standard deviation of $\log H_i$ match those of log(IVOL) in the empirical data. We set $\text{corr}[\, \log C_i, \log \sigma_i \,]$ to be -0.5 so that in the model $\text{corr}[\, \log V_i, \log H_i \,]$ matches the observed negative correlation between log(ME) and log(IVOL).

The distribution of factor exposure, B_i, has a mean of 1 and a standard deviation of

[15] The number of the firms traded on the NYSE, AMEX, and NASDAQ varies from about 500 in 1926, to about 3000 in 1967, to about 9000 in 1997, and to slightly less than 6000 in 2008.

[16] More specifically, in a sample of more than 62,000 household investors from a U.S. brokerage house, Goetzmann and Kumar (2008) show that more than 25% of the investor portfolios contain only one stock, over half of the investor portfolios contain no more than three stocks, and less than 10% the investor portfolios contain more than ten stocks.

0.5. The correlations between stock specific shocks, corr$[\,\varepsilon_{i1}, \varepsilon_{i2}\,]$, are assigned with a mean of 0 and and a standard deviation of 0.1. The model solutions indicate that the results are robust to changes in these parameters.[17]

As discussed earlier, in the empirical data, the Fama-MacBeth regressions of log(CSHR) on log(ME) suggest a roughly square-root dependence. Hence, we set $\lambda = 0.5$ so that the number of investors on average increases with the square root of C_i in our model.[18] The standard deviation σ_π, characterizing the variation in the number of investors orthogonal to C_i, is set to 1.3 so that the model matches the average R^2 of the regressions of log(CSHR) on log(ME).

Finally, the standard deviation of the macroeconomic factor, σ_F, set to 0.055, is intended for the model to match the monthly volatility of the aggregate stock market return. The monthly gross risk-free rate R_f is set to 1.003. The preference parameter $\delta = 0.7$, and for parsimony, is assumed to be the same for all investors. These two parameters are chosen to replicate the averages of the stock returns and the excess returns.

4.5 Model-implied results

The full empirical sample contains 1008 months, and we report the empirical results as time-series averages. Accordingly, we simulate the model economy for 1000 times, and solve for the equilibrium for each economy. From the solutions, we obtain directly an array of variables, in particular expected returns R_i, firm values V_i, and idiosyncratic volatilities H_i. We compute the averages of the model results over the 1000 equilibria.

We first compute the averages of key summary statistics over the 1000 equilibria and compare with those of the empirical data. We find that, on average, the mean of $\log V_i$ is 5.00, and the standard deviation is 1.75; the mean of $\log H_i$ is 2.15, and the standard

[17]We obtain very similar results if we change the standard deviation of B_i or specify the stock specific shocks as uncorrelated across firms.

[18]The model results are stronger if we set a smaller λ.

deviation is 0.65; the mean of $R_i - R_f$ is 0.90% per month; the correlation between $\log V_i$ and $\log H_i$ is -0.50. For the regressions of log number of investors on log size ($\log N_i$ on $\log V_i$), the average slope is 0.50, and the average R^2 is 0.31. The average market excess return is 0.60% per month, and the average market return volatility is 5.6% per month. These results confirm that the calibrated model well replicates the key features of the empirical data.

Next, we investigate the relations between size, idiosyncratic volatility, and expected return in the model, using the return regressions and portfolio sorting applied earlier to the empirical data. We report the averages of the results over the 1000 equilibria of the simulated economies and compare with the empirical results.

Table 10 presents the regressions of the expected excess returns of individual stocks. Regressing the expected excess return $R_i - R_f$ on size $\log V_i$ yields a negative coefficient. Regressing $R_i - R_f$ on idiosyncratic volatility $\log H_i$ yields a positive coefficient. When both variables are included, the coefficient for $\log V_i$ turns positive, while that for $\log H_i$ remains positive and become larger. These model-implied regression results well replicate the patterns observed in the real data regressions.

Table 11 presents the expected excess returns of the stock portfolios. In Panel A, stocks are sorted by size V into deciles of equal number of stocks,[19] and the equal-weighted expected returns in excess of R_f are reported. The results indicate higher expected returns for smaller stocks, consistent with the finding in the real data that small firms earn higher returns than large firms. Panel B reports the equal-weighted expected returns for the deciles sorted by idiosyncratic volatility H. The results indicate that expected return increases with idiosyncratic volatility. This is consistent with the finding in the real data that stocks with high idiosyncratic volatility earn high returns.

Finally, we apply the sequential sorting procedure: we sort the stocks first by idiosyncratic volatility H into deciles, then for each H decile, sort stocks by size V into ten portfolios.

[19] There is no model counterpart to NYSE stocks so we simply form portfolios of equal number of stocks.

To check the control for idiosyncratic volatility, Panel C reports the difference in median H between the largest and the smallest V portfolios within each H decile. For each of H deciles 1 to 9, the difference in median H is slightly negative and no more than 1% in magnitude. For H decile 10, the difference is about -11%. These results are quantitatively very close to the corresponding empirical results presented in Table 7. Therefore, in both the empirical data and the model equilibria, the control for idiosyncratic volatility is effective in idiosyncratic volatility deciles 1 to 9, but poor in decile 10.

Panel D presents the equal-weighted expected excess returns of the 100 H-then-V sorted portfolios. Within each of H deciles 1 to 9, the expected excess return increases with size — large stocks earn higher returns than small stocks. The only exception is in the highest H decile. Like the highest IVOL decile in the real data, the return spread between the highest and the lowest size portfolios is negative in the highest H decile.

Panel E of Table 11 presents the value-weighted expected excess return spreads. The results are qualitatively the same as those based on equal-weighted returns, though somewhat smaller in magnitudes. All these return patterns are similarly observed in the real data.

In our model, the CAPM beta varies independently from size and idiosyncratic volatility in the cross section. Therefore, the relations between return, size, and idiosyncratic volatility reported above are not driven by variations in the CAPM beta. For example, if we use the CAPM beta adjusted expected returns to compute the portfolio return spreads, the results are very close to those in Table 11. If we include the CAPM beta as a regressor in the return regressions, the coefficients on size and idiosyncratic volatility change very little.

Altogether, in the model equilibria, unconditionally there is a negative size-return relation, and a positive relation between idiosyncratic volatility and return. Most importantly, after controlling for idiosyncratic volatility, the relation between size and return becomes positive. The model is well capable of generating the qualitative patterns of the empirical results, while the quantitative magnitudes are smaller than those in the real data. Given the stylized nature of the model, its quantitative performance is still remarkable.

4.6 Counterfactual experiments

An advantage of the model is that it allows us to make counterfactual changes on the inputs to the model and examine the impact on the model results. In doing so, we can identify the underlying economic mechanisms that drive the particular model results.

Our model is based on three key empirical regularities — the decreasing size-scaled number of investors, the negative correlation between size and idiosyncratic volatility, and investor under-diversification. Thus, we conduct a few experiments and make counterfactual changes on each of the three features, one at a time. The results from these experiments are presented in Table 12. To facilitate comparison, the results for the baseline model are reproduced in line 0 of Table 12.

In the first experiment, we change λ to 1, so that the number of investors increases linearly with size, or equivalently, the size-scaled number of investors is the same for all stocks. Everything else remains the same. As reported in line 1 of Table 12, the size-return relation conditional on idiosyncratic volatility exhibits striking differences from the baseline model. In the baseline model (line 0), the relation is positive. When the number of investors is set to grow linearly with firm size, the relation becomes largely flat. In the return regression on both size ($\log V_i$) and idiosyncratic volatility ($\log H_i$), the slope on $\log V_i$ is essentially zero. In the portfolio return results, within each of H deciles 1 to 9 (where the control for H is effective), the return spreads between the largest and the smallest size portfolios are very close to zero.

In the second experiment, we change ρ to 0, so that size and idiosyncratic volatility vary independently across firms. Everything else remains the same, including $\lambda = 0.5$. The results are reported in line 2 of Table 12. Similar to the baseline model, this experiment also generates a positive size-return relation conditional on idiosyncratic volatility. Different from the baseline model, the relation between size and return is positive even without controlling for idiosyncratic volatility. As a matter of fact, since $\log V_i$ and $\log H_i$ are uncorrelated, the slope coefficients for them are essentially the same in the return regressions on $\log V_i$ and $\log H_i$ separately and jointly.

The results from the two experiments imply that the positive size-return relation is driven by the decreasing size-scaled number of investors, while the unconditional negative size-return relation is due to the negative correlation between size and idiosyncratic volatility. What is common across the baseline, the first experiment ($\lambda = 1$), and the second experiment ($\rho = 0$) is the increasing expected return in idiosyncratic volatility. This suggests that the positive relation between idiosyncratic volatility and return is driven by investor under-diversification.

We conduct two additional experiments to further investigate the role of investor under-diversification. Investors hold four stocks in the baseline model. In line 3, we let investors hold only three stocks; in line 4, they hold ten stocks. As the number of stocks increases from three to four and then to ten, investors become more diversified, and all the relations weaken in magnitude: the size-return relation conditional on idiosyncratic volatility becomes less positive; the relation between idiosyncratic volatility and return becomes less positive; the unconditional size-return relation becomes less negative. These results suggest a critical role of investor under-diversification in driving the relations between size, idiosyncratic volatility, and return.

To summarize, our results from the counterfactual experiments suggest the following economic intuition. First, because large stocks, relative to their size, attract fewer investors, they have to offer higher expected returns to induce their investors to allocate more of their wealth. This gives rise to the positive size-return relation. Second, when investors are under-diversified, they demand compensation for bearing idiosyncratic risk, resulting in the positive relation between idiosyncratic volatility and expected return. Finally, large (small) stocks tend to have low (high) idiosyncratic volatilities. This negative correlation between size and idiosyncratic volatility generates a negative link between size and return, and quantitatively, it more than offsets the positive link driven by the decreasing number of investors. As a result, we observe an unconditional negative relation.

In the first experiment ($\lambda = 1$), the number of investors is proportional to size. Relative to small firms, large firms do not need to offer higher expected returns to attract investor wealth. Hence, the size-return relation conditional on idiosyncratic volatility is flat. However, the negative link between size and return still exists due to the negative correlation between size and idiosyncratic volatility. As a result, small stocks exhibit high returns, and this is simply and wholly due to their large idiosyncratic volatilities.[20]

In the second experiment ($\rho = 0$), the correlation between size and idiosyncratic volatility becomes zero. This shuts down the negative link between size and return via idiosyncratic volatility. Meanwhile, $\lambda = 0.5$, the size-scaled number of investors decreases with size. This generates a positive relation between size and return, both with and without the control for idiosyncratic volatility.

4.7 Additional robustness checks

We study several additional variants of the model to check the robustness of the model implications.

In our baseline model, investors solve an unconstrained mean-variance portfolio optimization. The portfolio weights can be negative: investors can short stocks and borrow at the risk-free rate. In reality, investors often face constraints on shorting stocks and borrowing. We investigate two variants of the model that impose these restrictions. Line 5 in Table 12 presents the results when investors are not allowed to short stocks, and line 6 explores the consequences of imposing a constraint that investors cannot borrow more than 30% of

[20]The results in line 1 also indicate a negative expected return spread across the size portfolios in the highest H decile. This is due to the poor control for H which yields a negative H spread in this decile.

the net worth of their portfolios. With constraints, the size-return relation conditional on idiosyncratic volatility becomes more positive, and the unconditional size-return relation becomes somewhat less negative. In addition, the unconditional idiosyncratic volatility-return relation becomes less positive than that in the baseline model.

Another assumption we make for parsimony in the baseline model is that all investors have equal wealth. In a variant of the model, we divide the investors into two equal-numbered groups of 100,000 investors. Each investor in the first group is endowed with three times as much wealth as that of an investor in the second group. Consistent with the empirical evidence in Goetzmann and Kumar (2008) that larger stock portfolios by individual households are more diversified, each investor of the first group picks six stocks, while each investor of the second group picks two stocks. The results from this variant are very close to those of the baseline model. In yet another variant, we allocate 50,000 investors to the first, wealthier group, and the remaining 150,000 to the second group. The results are quantitatively stronger than those of the baseline model.

In addition, we increase the number of stocks to 4000, or the number of investors to 4×10^5. The model results of these two variants are very close to those of the baseline model.

4.8 Large and diversified funds

The baseline model focuses on under-diversified individual investors. This is consistent with the observation that such investors play a dominant role in the US stock market for a large part of our sample period. In recent decades, institutional investors are playing a more and more important role in the stock market. The total number of institutional investors as reported in the 13f institutional holding dataset has risen from 525 to 3100 during the period 1980–2008. The percentage of the total market capitalization held by institutional investors also increases from 31% to 68%, with a time-series average of 50%. The median (mean) number of stocks held by an institutional investor reduces from 125 to 70 (increases from 165 to 265) during the same period. The stock portfolios of typical institutions are large in value and contain many stocks, and thus are likely diversified. A growing number of individual investors hold stocks through these institutions and thus become diversified.

To explore the implications of this development, we consider a parsimonious setup with two groups of individual investors: the first group remain under-diversified, each holding four individual stocks; the second group hold stocks through 1000 large and diversified institutions, which we call mutual funds. As a result, our model accommodates diversified individual investors.[21]

The number of under-diversified individual investors is kept at 2×10^5. We also double the number of stocks to 4000. This is consistent with the trend of growing total number of stocks in the empirical data. Half of the total stock market wealth is held by under-diversified individual investors, and the other half under the management of mutual funds. Hence, in terms of the value of stock holdings, each mutual fund is as large as 200 under-diversified individual investors. Each mutual fund makes 50 stock picks. Again, for parsimony, we assume that mutual funds pick stocks with a probability following the square-root dependence on the cash flow. This is the same as that for under-diversified individual investors. On the other hand, the random variation terms in the probabilities representing the components orthogonal to firm size are independent between mutual funds and under-diversified individual investors.[22] Further, we assume that mutual funds are also mean-variance optimizing.

[21]In this parsimonious setup, under-diversified individual investors do not invest in mutual funds. This simplifies the numerical solution of the equilibrium. Specifically, in this setup, for both under-diversified individual investors and mutual funds, portfolio weights depend on the expected returns and the covariances of the stocks. If under-diversified individual investors also invest in mutual funds, then these investors' portfolio weights also depend on the expected returns and the covariances of mutual funds, which, in turn, depend on mutual funds' portfolio weights. We find that this additional layer of dependence substantially slows down the convergence of the numerical solution.

[22]The results are similar if the random variation terms are the same for mutual funds and under-diversified individual investors.

Since institutional investors are usually restricted from shorting or borrowing,[23] we impose the restrictions of no shorting and borrowing on the mutual funds in our model. The model solution indicates that mutual funds on average put positive weights in about 23 different stocks.

Overall, we make a number of simplifying assumptions on the behavior of the mutual funds in our model. Our focus is how the results change when both small, under-diversified individual investors and large institutions representing diversified individual investors participate in the stock market. Line 7 in Table 12 presents the results. Compared to the baseline model, the qualitative patterns remain the same under this setup. On the other hand, all the relations become smaller in magnitude: the unconditional size effect is less negative; the unconditional idiosyncratic volatility effect is less positive; the return spreads between the largest and the smallest V portfolios for H deciles 1 to 9 are less positive than those of the baseline model. These changes from the baseline model due to the participation of mutual funds are anticipated, as they confirm investor under-diversification as the economic driving force of the relations between size, idiosyncratic volatility, and return. In addition, these changes from the baseline model due to the participation of mutual funds are broadly consistent with the variations in the empirical results from the early period 1926:07–1967:12 to the later period 1968:01–2009:12. Institutional investors play a more important role in the stock market in the later period.

5 Concluding remarks

In this study, we document robust empirical evidence that, after controlling for idiosyncratic volatility, there is a positive relation between size and return. We explain this positive relation, and reconcile it with the widely documented negative relation without the control for

[23]Under federal law, mutual funds cannot take on debt of more than a third of their assets.

idiosyncratic volatility. We demonstrate the economic intuition in a parsimonious equilibrium model. Specifically, some individual investors are under-diversified, and large stocks, relative to their size, are held by fewer investors. To clear the markets, large stocks have to offer higher expected returns to induce the investors to allocate more of their wealth, giving rise to the positive size-return relation. The traditional negative relation is the result of the failure to control for idiosyncratic volatility. This generates a downward omitted variable bias because idiosyncratic volatility is negatively related to size and, in the presence of investor under-diversification, positively related to return.

The model in our study reveals important insights that do not readily transpire in the factor pricing framework. Indeed, the economic intuition underlying the relations between size, idiosyncratic volatility, and return is difficult to explain with factor pricing models. First, the positive size-return relation in our model is not driven by variations in either systematic risk or idiosyncratic risk.[24] Second, idiosyncratic volatility is associated with high returns in our model because it contributes to the risk of under-diversified portfolios. As investors hold a small number of different stocks, the same stock contributes varying amount of risk to different investor portfolios. This is difficult to fit into factor pricing models, which assume that a stock has the same risk to all investors.

Our model is based on three empirical regularities. We take them as inputs to our model. In particular, following Levy (1978) and Merton (1987), we assign under-diversified portfolios to investors and investigate the asset pricing implications. It would be interesting to incorporate potential mechanisms into the model so that these empirical regularities (in particular, investor under-diversification and the decreasing size-scaled number of investors) could arise endogenously. Such an extension is computationally challenging and merits a separate study in the future. In our current paper, we have investigated a few model variants

[24]In the baseline model, size and the CAPM beta are uncorrelated. In one of the counterfactual experiments, we further set zero correlation between size and idiosyncratic volatility. The positive size-return relation arises in both scenarios.

— we change the number of stocks in investor portfolios, divide investors into two groups holding different numbers of stocks, or introduce large and diversified mutual funds — and find similar results. While these variants do not incorporate the endogeneity directly, they suggest that the model results are robust to different scenarios of investor portfolio selections.

Following the classical CAPM literature, we employ a two-period setup for our model. This static setup precludes time-varying returns and volatilities. A dynamic setup could lead to new and interesting implications.[25] To sharpen the focus of this paper, we leave the extension to future research.

Appendix: Omitted variable bias

For expositional clarity, we suppress the time and firm subscripts. In

$$\text{XRET} = \text{const} + b_1 \log(\text{ME}) + b_2 \log(\text{IVOL}) + u, \tag{26}$$

the empirical results suggest that

$$b_1 > 0, \quad b_2 > 0, \quad \text{corr}[\log(\text{ME}), \log(\text{IVOL})] < 0. \tag{27}$$

Then, regressing XRET on log(ME) only, with log(IVOL) as the omitted variable,

$$\text{XRET} = \text{const} + b_1' \log(\text{ME}) + u_1, \tag{28}$$

we obtain a slope of

$$b_1' = \frac{\text{cov}[\text{XRET}, \log(\text{ME})]}{\text{var}[\log(\text{ME})]} \tag{29}$$

$$= \frac{b_1 \text{var}[\log(\text{ME})] + b_2 \text{cov}[\log(\text{IVOL}), \log(\text{ME})]}{\text{var}[\log(\text{ME})]} \tag{30}$$

$$= b_1 + b_2 \text{ corr}[\log(\text{ME}), \log(\text{IVOL})] \frac{\text{std}[\log(\text{IVOL})]}{\text{std}[\log(\text{ME})]}. \tag{31}$$

[25] For example, Shapiro (2002) studies investor recognition in an intertemporal setting and reveals new insights.

The two terms in the above suggest two channels that drive the unconditional relation between size and return. The first channel is the positive size-return relation ($b_1 > 0$). In the second channel, size is negatively correlated with idiosyncratic volatility, which, in turn, is positively associated with return ($b_2 > 0$). The second channel represents the omitted variable bias.

Regressing XRET on log(IVOL) only, with log(ME) as the omitted variable,

$$\mathrm{XRET} = \mathrm{const} + b_2' \log(\mathrm{IVOL}) + u_2, \tag{32}$$

we obtain a slope of

$$b_2' = \frac{\mathrm{cov}[\mathrm{XRET}, \log(\mathrm{IVOL})\,]}{\mathrm{var}[\,\log(\mathrm{IVOL})\,]} \tag{33}$$

$$= \frac{b_1\,\mathrm{cov}[\,\log(\mathrm{ME}), \log(\mathrm{IVOL})\,] + b_2\,\mathrm{var}[\,\log(\mathrm{IVOL})\,]}{\mathrm{var}[\,\log(\mathrm{IVOL})\,]} \tag{34}$$

$$= b_2 + b_1\ \mathrm{corr}[\,\log(\mathrm{ME}), \log(\mathrm{IVOL})\,]\frac{\mathrm{std}[\,\log(\mathrm{ME})\,]}{\mathrm{std}[\,\log(\mathrm{IVOL})\,]}. \tag{35}$$

There are also two channels driving the unconditional relation between idiosyncratic volatility and return. The first channel is the positive relation ($b_2 > 0$). In the second channel, idiosyncratic volatility is negatively related to size, which, in turn, is positively associated with return ($b_1 > 0$). The second channel represents the omitted variable bias.

The negative corr[log(ME), log(IVOL)] implies that the second channels counteract the first channels — i.e., the biases are downward. Hence, $b_1' < b_1$ and $b_2' < b_2$. Further, $b_1' < 0$ implies $b_2' > 0$. More specifically, $b_1' < 0$ suggests that

$$\frac{b_1}{b_2} < \Big(- \mathrm{corr}[\,\log(\mathrm{ME}), \log(\mathrm{IVOL})\,]\Big)\frac{\mathrm{std}[\,\log(\mathrm{IVOL})\,]}{\mathrm{std}[\,\log(\mathrm{ME})\,]}. \tag{36}$$

This implies that, as the magnitude of the correlation coefficient is smaller than 1,

$$\frac{b_1}{b_2} < \frac{1}{\Big(- \mathrm{corr}[\,\log(\mathrm{ME}), \log(\mathrm{IVOL})\,]\Big)}\frac{\mathrm{std}[\,\log(\mathrm{IVOL})\,]}{\mathrm{std}[\,\log(\mathrm{ME})\,]}, \tag{37}$$

which implies $b_2' > 0$.

References

Amihud, Y., H. Mendelson, and J. Uno (1999). Number of shareholders and stock prices: Evidence from Japan. *Journal of Finance 54*, 1169–1184.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006). The cross-section of volatility and expected returns. *Journal of Finance 61*, 259–299.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics 9*, 3–18.

Barber, B. M. and T. Odean (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *Journal of Finance 55*, 773–806.

Barberis, N. C. and M. Huang (2008). Stocks as lotteries: The implications of probability weighting for security prices. *American Economic Review 98*, 2066–2100.

Berk, J., R. C. Green, and V. Naik (1999). Optimal investment, growth options, and security returns. *Journal of Finance 54*, 1553–1607.

Berk, J. B. (1995). A critique of size-related anomalies. *Review of Financial Studies 8*, 275–286.

Bloomfield, T., R. W. Leftwich, and J. B. Long (1977). Portfolio strategies and performance. *Journal of Financial Economics 5*, 201–218.

Blume, M. and I. Friend (1975). The asset structure of individual portfolios and some implications for utility functions. *Journal of Finance 30*, 585–603.

Bodnaruk, A. and P. Ostberg (2009). Does investor recognition predict returns? *Journal of Financial Economics 91*, 208–226.

Brennan, M. J. (1975). The optimal number of securities in a risky asset portfolio when there are fixed costs of transacting: Theory and some empirical results. *Journal of Financial and Quantitative Analysis 10*, 483–496.

Calvet, L. E., J. Y. Campbell, and P. Sodini (2007). Down or out: Assessing the welfare costs of household investment mistakes. *Journal of Political Economy 115*, 707–747.

Campbell, J. Y. (2006). Household finance. *Journal of Finance 61*, 1553–1604.

Chan, K. C. and N.-F. Chen (1991). Structural and return characteristics of small and large firms. *Journal of Finance 46*, 529–554.

Chordia, T., A. Subrahmanyam, and V. R. Anshuman (2001). Trading activity and expected stock returns. *Journal of Finance 59*, 3–32.

Cohen, L. (2009). Loyalty-based portfolio choice. *Review of Financial Studies 22*, 1213–1245.

Dimson, E. (1979). Risk measurement when shares are subject infrequent trading. *Journal of Financial Economics 7*, 197–226.

Dyl, E. A. and W. B. Elliott (2006). The share price puzzle. *Journal of Business 79*, 2045–2066.

Fama, E. F. and K. R. French (1992). The cross-section of expected stock returns. *Journal of Finance 47*, 427–465.

Foerster, S. R. and G. A. Karolyi (1999). The effects of market segmentation and investor recognition on asset prices: Evidence from foreign stocks listing in the United States. *Journal of Finance 54*, 981–1013.

Friend, I., R. Westerfield, and M. Granito (1978). New evidence on the capital asset pricing model. *Journal of Finance 33*, 903–917.

Fu, F. (2009). Idiosyncratic risk and the cross-section of expected stock returns. *Journal of Financial Economics 91*, 24–37.

Gervais, S., R. Kaniel, and D. Mingelgrin (2001). The high-volume return premium. *Journal of Finance 56*, 877–919.

Goetzmann, W. N. and A. Kumar (2008). Equity portfolio diversification. *Review of Finance 12*, 433–463.

Gomes, J. F., L. Kogan, and L. Zhang (2003). Equilibrium cross section of returns. *Journal of Political Economy 111*, 693–732.

Harvey, C. R. and A. Siddique (2000). Conditional skewness in asset pricing tests. *Journal of Finance 55*, 1263–1295.

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *Journal of Finance 45*, 881–898.

Jegadeesh, N. and S. Titman (2001). Profitability of momentum strategies: An evaluation of alternative explanations. *Journal of Finance 56*, 699–720.

Kadlec, G. B. and J. J. McConnell (1994). The effect of market segmentation and illiquidity on asset prices: Evidence from exchange listings. *Journal of Finance 49*, 611–636.

Kelly, M. (1995). All their eggs in one basket: Portfolio diversification of US households. *Journal of Economic Behavior and Organization 27*, 87–96.

Kraus, A. and R. Litzenberger (1976). Skewness preference and the valuation of risk assets. *Journal of Finance 31*, 1085–1100.

Levy, H. (1978). Equilibrium in an imperfect market: A constraint on the number of securities in the portfolio. *American Economic Review 68*, 643–658.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics 47*, 13–37.

Liu, H. (2009). Portfolio insurance and underdiversification. Unpublished working paper. Washington University in St. Louis.

Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information. *Journal of Finance 42*, 483–510.

Odean, T. (1999). Do investors trade too much? *American Economic Review 89*, 1279–1298.

Polkovnichenko, V. (2005). Household portfolio diversification: A case for rank dependent preferences. *Review of Financial Studies 18*, 1467–1502.

Reinganum, M. R. (1981). Misspecification of capital asset pricing: Empirical anomalies based on earnings' yields and market values. *Journal of Financial Economics 9*, 19–46.

Schwert, G. W. (1983). Size and stock returns, and other empirical regularities. *Journal of Financial Economics 12*, 3–12.

Schwert, G. W. (2003). Anomalies and market efficiency. In G. M. Constantinides, M. Harris, and R. M. Stulz (Eds.), *Handbook of the Economics of Finance*, pp. 939–974. Amsterdam: Elsevier.

Shapiro, A. (2002). The investor recognition hypothesis in a dynamic general equilibrium: Theory and evidence. *Review of Financial Studies 15*, 97–141.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance 19*, 425–442.

Shefrin, H. and M. Statman (2000). Behavioral portfolio theory. *Journal of Financial and Quantitative Analysis 35*, 127–151.

Shumway, T. and V. Warther (1999). The delisting bias in CRSP's Nasdaq data and its implications for the size effect. *Journal of Finance 55*, 2361–2379.

Van Nieuwerburgh, S. and L. Veldkamp (2010). Information acquisition and under-diversification. *Review of Economic Studies 77*, 779–805.

Table 1: Time-series averages of cross-sectional descriptive statistics of the sample

This table reports the time-series averages of the descriptive statistics of the empirical data variables. We use log() to denote natural logarithm.

In Panel A, we begin with all stocks traded on the NYSE, AMEX, and NASDAQ during the period of January 1926–December 2009, and require that monthly return (RET), end-of-previous-month market capitalization (ME), and idiosyncratic volatility (IVOL) are available. There are 3,520,244 firm-month observations in total. RET is the monthly raw return in percentage. XRET is the monthly return in excess of the one-month T-bill rate. ME is in millions of dollars, and adjusted by the Consumer Price Index to the dollars of December 2000. IVOL is estimated each month for each stock, as described in Section 2. To mitigate the influence of extreme values, we delete monthly returns over 300% and winsorize the top and bottom 0.5% of IVOL estimates in each month. We evaluate the descriptive statistics for each month, and report the time-series averages.

In Panel B, CSHR is the number of common shareholders in thousands. Compustat reports this data item since fiscal year 1975. The statistics for this variable is based on 201,062 firm-year observations for the period 1975–2009. We evaluate the descriptive statistics for each year, and report the time-series averages.

Variable	Mean	Std Dev	Skewness	Quartile 1	Median	Quartile 3
Panel A: Monthly data, 1926:01–2009:12						
RET (%)	1.23	12.55	1.69	-5.28	0.25	6.30
XRET (%)	0.93	12.55	1.69	-5.58	-0.05	6.00
ME ($ million)	978.77	4488.89	15.17	52.83	154.46	562.34
IVOL (%)	11.39	8.60	2.23	5.87	8.99	14.02
log(ME)	5.00	1.76	0.31	3.76	4.87	6.13
log(IVOL)	2.15	0.63	0.14	1.72	2.14	2.56
Panel B: Annual data, 1975–2009						
CSHR (thousand)	15.61	323.14	35.36	0.62	1.55	4.80
log(CSHR)	0.53	1.72	0.41	-0.64	0.36	1.53

Table 2: Time-series averages of cross-sectional simple correlations

The sample is monthly for the period of January 1926–December 2009. We use log() to denote natural logarithm.

In each month, we compute the correlations between monthly return (RET), end-of-previous-month market capitalization (ME), and idiosyncratic volatility (IVOL). This table presents the time-series averages of the correlations. The t-statistics in the parentheses are the time-series averages of the correlations divided by the corresponding time-series standard errors.

	log(ME)	log(IVOL)
RET	-0.01 (-2.01)	0.09 (12.55)
log(ME)		-0.52 (-113.1)

Table 3: Fama-MacBeth regressions of excess returns

The sample is monthly for the period of January 1926–December 2009. We use log() to denote natural logarithm.

In each month we run cross-sectional regressions. XRET is the percentage monthly return in excess of the one-month T-bill rate, ME is end-of-previous-month market capitalization, and IVOL is idiosyncratic volatility.

We report the results for the full sample period of January 1926–December 2009 and separately for two subperiods: the early period 1926:01–1967:12 and the later period 1968:01–2009:12. For each sample period, we report the time-series averages of the slope coefficients and the R^2 values. The t-statistics in the parentheses are the time-series averages of the slopes divided by the corresponding time-series standard errors. For the full sample period, there are a total of 1007 monthly regressions, with an average of 3495 firms per month. For the early subperiod, there are a total of 503 monthly regressions, with an average of 1017 firms per month. For the later subperiod, there are a total of 504 monthly regressions, with an average of 5966 firms per month.

Time-series averages		
Slope on log(ME)	Slope on log(IVOL)	R^2 (%)
Full sample: 1926:01–2009:12		
-0.18 (-4.39)		1.95
	2.14 (12.48)	5.48
0.33 (12.85)	2.66 (15.29)	6.56
Early period: 1926:01–1967:12		
-0.24 (-3.40)		2.65
	2.44 (9.47)	6.52
0.46 (12.32)	3.22 (12.31)	7.91
Later period: 1968:01–2009:12		
-0.12 (-2.85)		1.25
	1.83 (8.15)	4.45
0.20 (5.85)	2.09 (9.26)	5.21

Table 4: Fama-MacBeth regressions of excess returns with additional control variables in the period of 1968:01–2009:12

The sample is monthly for the period of 1968:01–2009:12. We use log() to denote natural logarithm.

In each month we run cross-sectional regressions. XRET is the percentage monthly return in excess of the one-month T-bill rate, ME is end-of-previous-month market capitalization, IVOL is idiosyncratic volatility, B/M is book-to-market equity, TURN is the average turnover in the past 36 months, CVTURN is the coefficient variation of the past 36-month turnovers, RET_{t-1} is the preceding month return, and $\mathrm{RET}_{t-7,t-2}$ is the compound gross return of the previous 6 months (skipping the preceding month).

We report the time-series averages of the slope coefficients and the R^2 values. The t-statistics in the parentheses are the time-series averages of the slopes divided by the corresponding time-series standard errors.

	Time-series averages							
Dependent variable	Slope on log(ME)	Slope on log(IVOL)	Slope on log(B/M)	Slope on log(TURN)	Slope on log(CVTURN)	Slope on RET_{t-1}	Slope on $\mathrm{RET}_{t-7,t-2}$	R^2 (%)
XRET	-0.22 (-5.65)		0.14 (2.78)	-0.01 (-0.15)	-0.38 (-4.70)	-0.06 (-12.65)	0.56 (2.19)	7.67
XRET	0.19 (6.38)	2.56 (13.10)	0.36 (7.90)	-0.45 (-7.52)	-0.63 (-8.34)	-0.06 (-15.00)	0.72 (3.16)	10.66
log(100+XRET)*100	0.16 (5.65)	1.26 (6.72)	0.39 (8.93)	-0.51 (-8.45)	-0.63 (-8.49)	-0.06 (-15.98)	0.80 (3.74)	10.57

Table 5: Time-series averages of excess returns for portfolios sorted by size

The sample is monthly for the period of January 1926–December 2009. In each month we sort stocks into 10 portfolios by end-of-previous-month market capitalization (ME). The breaking points are based on the ME of NYSE stocks only.

In each month, for each ME portfolio, we compute the equal- and value-weighted returns in excess of the one-month T-bill rate. We report the time-series averages of the portfolio excess returns (XRET) for the full sample period of January 1926–December 2009 and separately for three subperiods: the early period 1926:01–1967:12, the later period 1968:01–2009:12, and the latest period 1982:01–2009:12, which is after the documentation of the size effect in Banz (1981) and Reinganum (1981). FF alphas are the intercepts estimated from the time-series regressions of the portfolio return spreads on the Fama-French three factors.

ME XRET (%)	1 Small	2	3	4	5	6	7	8	9	10 Large	ME 10-1 Spread (t-stat)	FF alpha (t-stat)
Full sample: 1926:01–2009:12												
EW	1.62	0.99	0.83	0.85	0.84	0.82	0.76	0.75	0.71	0.58	-1.04 (-4.12)	-0.33 (-2.36)
VW	1.28	0.98	0.83	0.85	0.83	0.82	0.76	0.75	0.71	0.57	-0.72 (-2.92)	0.02 (0.20)
Early period: 1926:01–1967:12												
EW	2.46	1.49	1.20	1.18	1.10	1.10	1.00	0.97	0.91	0.75	-1.71 (-3.80)	-0.87 (-4.23)
VW	2.10	1.47	1.21	1.17	1.08	1.09	1.01	0.97	0.90	0.77	-1.33 (-3.03)	-0.45 (-2.51)
Later period: 1968:01–2009:12												
EW	0.78	0.50	0.45	0.53	0.58	0.55	0.51	0.54	0.51	0.41	-0.38 (-1.64)	-0.05 (-0.31)
VW	0.47	0.49	0.45	0.53	0.58	0.55	0.50	0.54	0.51	0.36	-0.11 (-0.49)	0.28 (2.18)
Latest period: 1982:01–2009:12												
EW	0.76	0.54	0.56	0.62	0.67	0.62	0.67	0.67	0.66	0.63	-0.13 (-0.49)	-0.04 (-0.17)
VW	0.42	0.54	0.56	0.62	0.67	0.63	0.66	0.67	0.67	0.57	0.14 (0.60)	0.32 (1.88)

Table 6: Time-series averages of excess returns for portfolios sorted by idiosyncratic volatility

The sample is monthly for the period of January 1926–December 2009. In each month we sort stocks into 10 portfolios by idiosyncratic volatility (IVOL). The breaking points are based on the IVOL of NYSE stocks only.

In each month, for each IVOL portfolio, we compute the equal- and value-weighted returns in excess of the one-month T-bill rate. We report the time-series averages of the portfolio excess returns (XRET) for the full sample period of January 1926–December 2009 and separately for two subperiods: the early period 1926:01–1967:12 and the later period 1968:01–2009:12. FF alphas are the intercepts estimated from the time-series regressions of the portfolio return spreads on the Fama-French three factors.

IVOL XRET (%)	1 Low	2	3	4	5	6	7	8	9	10 High	IVOL 10-1 Spread (t-stat)	FF alpha (t-stat)
					Full sample: 1926:01–2009:12							
EW	0.02	0.04	0.04	0.10	0.22	0.30	0.52	0.74	1.22	3.80	3.78 (10.81)	2.67 (14.14)
VW	0.32	0.45	0.51	0.59	0.84	0.98	1.18	1.59	1.93	2.84	2.52 (7.83)	1.83 (7.39)
					Early period: 1926:01–1967:12							
EW	0.15	0.16	0.15	0.26	0.54	0.67	1.04	1.46	2.23	4.92	4.76 (8.47)	3.49 (13.31)
VW	0.38	0.52	0.59	0.75	1.10	1.36	1.72	2.45	3.20	4.61	4.22 (9.08)	3.54 (9.24)
					Later period: 1968:01–2009:12							
EW	-0.11	-0.08	-0.08	-0.06	-0.09	-0.08	-0.01	0.03	0.21	2.69	2.80 (6.78)	2.10 (8.94)
VW	0.25	0.37	0.42	0.44	0.57	0.59	0.65	0.72	0.66	1.07	0.82 (1.89)	0.25 (0.92)

Table 7: Characteristics of IVOL-then-ME sorted portfolios

The sample is monthly for the period of January 1926–December 2009.

In each month we first sort stocks into 10 portfolios by idiosyncratic volatility (IVOL) of the current month, and then sort the stocks in each IVOL decile into 10 portfolios by end-of-previous-month market capitalization (ME). Both steps of sorting are based on the breaking points of NYSE stocks only.

In each month, for each IVOL*ME portfolio, we compute the median ME and the median IVOL of the stocks in the portfolio. We report the time-series averages of the portfolio characteristics.

ME	1	2	3	4	5	6	7	8	9	10	10-1
IVOL	Small									Large	Spread
					Panel A: Time-series averages of median ME ($ millions)						
1 Low	64.21	165.83	282.03	435.13	681.18	1047.50	1595.73	2472.88	4277.78	13326.83	13262.61
2	57.32	145.88	238.00	362.74	552.46	859.65	1337.07	2116.78	3767.39	10502.64	10445.32
3	52.23	131.60	222.00	348.06	522.34	789.48	1205.56	1939.82	3526.18	9335.60	9283.37
4	47.21	123.32	205.88	314.38	465.50	689.06	1048.80	1676.94	3091.49	8180.19	8132.98
5	42.59	112.11	184.09	275.58	398.64	578.96	866.79	1377.51	2560.71	7016.71	6974.12
6	37.81	98.39	159.26	235.73	337.71	483.77	711.00	1123.80	2058.26	5812.28	5774.47
7	32.49	84.30	134.14	196.03	276.13	390.18	567.42	885.44	1599.26	4487.97	4455.48
8	27.61	70.03	110.26	158.97	221.91	309.30	441.67	676.41	1208.01	3324.69	3297.09
9	21.47	53.43	83.19	119.59	165.42	227.47	319.87	480.96	839.15	2280.49	2259.02
10 High	10.66	27.33	43.48	63.42	89.97	125.97	178.64	268.17	460.09	1250.52	1239.87
					Panel B: Time-series averages of median IVOL (%)						
1 Low	3.07	3.11	3.13	3.14	3.15	3.15	3.15	3.13	3.10	3.04	-0.03
2	4.38	4.38	4.38	4.36	4.36	4.37	4.37	4.36	4.33	4.32	-0.06
3	5.30	5.30	5.28	5.29	5.28	5.28	5.27	5.27	5.25	5.24	-0.05
4	6.18	6.18	6.17	6.17	6.16	6.15	6.15	6.14	6.14	6.13	-0.05
5	7.12	7.12	7.11	7.09	7.10	7.08	7.08	7.07	7.07	7.04	-0.08
6	8.21	8.18	8.18	8.16	8.17	8.16	8.15	8.15	8.12	8.09	-0.12
7	9.56	9.52	9.52	9.49	9.47	9.47	9.47	9.44	9.43	9.41	-0.15
8	11.41	11.35	11.33	11.31	11.28	11.24	11.22	11.20	11.15	11.15	-0.27
9	14.56	14.44	14.34	14.25	14.20	14.18	14.04	14.03	13.96	13.92	-0.64
10 High	30.13	24.71	23.23	22.50	21.68	21.31	20.76	20.44	20.18	20.11	-10.02

Table 8: Time-series averages of the excess returns of IVOL-then-ME sorted portfolios

The sample is monthly for the period of January 1926–December 2009.

In each month we first sort stocks into 10 portfolios by idiosyncratic volatility (IVOL) of the current month, and then sort the stocks in each IVOL decile into 10 portfolios by end-of-previous-month market capitalization (ME). Both steps of sorting are based on the breaking points of NYSE stocks only.

In each month, for each IVOL*ME portfolio, we compute the equal- and value-weighted returns in excess of the one-month T-bill rate. Panel A reports the time-series averages of the portfolio excess returns (XRET) for the full sample period of January 1926–December 2009. FF alphas are the intercepts estimated from the time-series regressions of the portfolio return spreads on the Fama-French three factors. Panel B reports the time-series averages of the excess return spreads between the largest and the smallest ME portfolios within each of the 10 IVOL deciles for two subperiods: the early period 1926:01–1967:12 and the later period 1968:01–2009:12.

Table 8: continued.

Panel A: Time-series averages of portfolio excess returns

ME IVOL	1 Small	2	3	4	5	6	7	8	9	10 Large	ME 10-1 Spread (t-stat)	FF alpha (t-stat)
Full sample: 1926:01–2009:12. XRET, EW (%)												
1 Low	-0.28	-0.11	-0.03	0.06	0.11	0.15	0.25	0.24	0.22	0.35	0.64 (6.03)	0.39 (5.67)
2	-0.48	-0.15	-0.01	0.07	0.06	0.27	0.35	0.33	0.43	0.49	0.97 (8.06)	0.75 (8.68)
3	-0.68	-0.22	0.03	0.12	0.17	0.29	0.35	0.40	0.55	0.52	1.21 (9.28)	0.99 (10.36)
4	-0.68	-0.18	-0.04	0.14	0.29	0.33	0.43	0.72	0.69	0.69	1.37 (10.48)	1.24 (12.86)
5	-0.78	-0.03	0.08	0.41	0.48	0.56	0.57	0.78	0.95	0.99	1.77 (11.86)	1.72 (14.37)
6	-0.80	-0.14	0.21	0.43	0.52	0.67	0.94	0.96	1.03	1.19	1.99 (12.66)	1.80 (13.71)
7	-0.57	0.00	0.40	0.76	0.77	1.04	1.16	1.30	1.43	1.46	2.02 (11.93)	2.08 (13.78)
8	-0.60	0.13	0.51	0.84	1.04	1.51	1.64	1.74	1.84	1.84	2.45 (13.15)	2.54 (15.62)
9	-0.14	0.87	1.14	1.31	1.77	1.91	1.84	2.15	2.36	2.46	2.60 (10.77)	2.77 (12.78)
10 High	4.95	3.76	3.43	3.74	3.48	3.33	3.29	3.32	3.21	3.04	-1.91 (-4.38)	-1.38 (-3.42)
Full sample: 1926:01–2009:12. XRET, VW (%)												
1 Low	-0.23	-0.09	-0.03	0.06	0.12	0.15	0.25	0.25	0.21	0.40	0.63 (6.20)	0.44 (6.17)
2	-0.38	-0.14	-0.01	0.07	0.07	0.27	0.36	0.33	0.43	0.55	0.93 (7.59)	0.74 (8.01)
3	-0.62	-0.22	0.03	0.13	0.18	0.29	0.35	0.41	0.54	0.61	1.23 (9.25)	1.05 (9.98)
4	-0.55	-0.18	-0.03	0.14	0.29	0.33	0.44	0.73	0.70	0.68	1.23 (9.19)	1.20 (11.55)
5	-0.65	0.00	0.08	0.41	0.48	0.58	0.58	0.77	0.95	0.99	1.63 (9.78)	1.67 (11.64)
6	-0.63	-0.12	0.22	0.44	0.52	0.66	0.94	0.97	1.03	1.15	1.77 (10.74)	1.71 (11.74)
7	-0.48	0.03	0.41	0.78	0.79	1.04	1.16	1.31	1.44	1.34	1.82 (8.65)	2.00 (10.54)
8	-0.43	0.15	0.51	0.85	1.06	1.52	1.64	1.73	1.83	1.75	2.19 (10.79)	2.34 (12.78)
9	0.04	0.90	1.15	1.32	1.77	1.91	1.81	2.18	2.36	2.13	2.09 (7.74)	2.43 (10.03)
10 High	4.36	3.76	3.47	3.74	3.48	3.35	3.28	3.31	3.22	2.56	-1.80 (-3.84)	-1.22 (-2.84)

Table 8: continued.

Panel B: Time-series averages of excess return spreads between ME 10 and 1 portfolios within each IVOL decile

IVOL	1	2	3	4	5	6	7	8	9	10
ME 10-1	Low									High
				Early period: 1926:01–1967:12. XRET (%)						
EW	0.54	0.89	1.18	1.43	1.84	2.30	2.23	2.97	3.30	-0.51
(t-stat)	(3.10)	(4.49)	(5.45)	(6.73)	(7.32)	(8.63)	(7.92)	(9.50)	(7.85)	(-0.67)
VW	0.60	0.95	1.39	1.45	1.90	2.22	2.26	2.96	3.20	-0.72
(t-stat)	(3.75)	(4.70)	(6.36)	(6.70)	(6.57)	(8.09)	(6.17)	(8.76)	(6.77)	(-0.87)
				Later period: 1968:01–2009:12. XRET (%)						
EW	0.73	1.04	1.24	1.30	1.70	1.68	1.81	1.92	1.90	-3.31
(t-stat)	(6.16)	(7.72)	(8.50)	(8.61)	(10.49)	(10.14)	(9.65)	(9.64)	(8.14)	(-8.28)
VW	0.67	0.92	1.07	1.02	1.37	1.33	1.37	1.42	0.98	-2.87
(t-stat)	(5.23)	(6.51)	(7.07)	(6.41)	(8.16)	(7.28)	(6.73)	(6.46)	(3.89)	(-6.59)

Table 9: Number of shareholders and market capitalization

The sample is annual for the period of 1975–2009. We use log() to denote natural logarithm.

In each year, we compute the correlation between log fiscal-year-end market capitalization, log(ME), and log number of common shareholders, log(CSHR). Panel A reports the time-series average of the correlations. The t-statistic in the parentheses is the time-series average of the correlations divided by the corresponding time-series standard error.

In each year, we regress log(CSHR) on log(ME). Panel B reports the time-series averages of the slope coefficients and the R^2 values from these regressions. The t-statistic in the parentheses is the time-series average of the slopes divided by the corresponding time-series standard error. There are a total of 35 yearly regressions, with an average of 5743 firms per year.

Panel A: Time-series average of cross-sectional simple correlations (t-stat)

	log(ME)
log(CSHR)	0.57 (27.0)

Panel B: Fama-MacBeth regressions

Time-series averages	
Slope on log(ME)	R^2 (%)
0.42 (30.1)	33.9

Table 10: Regressions of expected excess returns in the baseline model

This table reports the regressions of the expected excess returns for individual stocks in the baseline model. We use log to denote natural logarithm. The independent variable is the expected excess return $R_i - R_f$ in percentage. The regressors are size $\log V_i$ and idiosyncratic volatility $\log H_i$. Here, i is the firm subscript. The coefficients (t-statistics) and R^2 are averages across 1000 simulations.

Slope on $\log V_i$	Slope on $\log H_i$	$R^2(\%)$
-0.092 (-4.53)		1.04
	0.922 (18.6)	14.9
0.112 (5.25)	1.072 (18.9)	16.0

Table 11: Portfolio expected excess returns in the baseline model

This table reports the portfolio sorting results for the baseline model. All results are averages across 1000 simulations.

In Panels A, B, and D, the portfolio expected excess returns are equal-weighted and in excess of the risk-free rate R_f. In Panel A, stocks are sorted by size V into deciles. In Panel B, stocks are sorted by idiosyncratic volatility H into deciles. In Panels C and D, stocks are first sorted into H deciles, and then, within each H decile, sorted into V deciles. Panel C presents the difference in median H between the largest and the smallest V portfolios within each H decile. Panel D presents the equal-weighted expected excess returns of the 100 H-then-V sorted portfolios. Panel E presents the value-weighted expected excess return spreads.

Panel A: V portfolio expected excess returns (%)

V	1 Small	2	3	4	5	6	7	8	9	10 Large	V 10-1 Spread
EW	1.21	0.99	0.87	0.78	0.70	0.64	0.59	0.56	0.54	0.53	-0.68

Panel B: H portfolio expected excess returns (%)

H	1 Low	2	3	4	5	6	7	8	9	10 High	H 10-1 Spread
EW	0.19	0.22	0.26	0.32	0.43	0.52	0.70	0.98	1.51	2.85	2.66

Table 11: continued

Panel C: Difference in median H of $V10$ and $V1$ (%)

H	1 Low	2	3	4	5	6	7	8	9	10 High
	-0.04	-0.05	-0.07	-0.10	-0.13	-0.17	-0.25	-0.58	-0.98	-11.03

Panel D: H*V portfolio expected excess returns, EW (%)

V / H	1 Small	2	3	4	5	6	7	8	9	10 Large	V 10-1 Spread
1 Low	0.17	0.17	0.18	0.18	0.19	0.19	0.20	0.21	0.22	0.24	0.07
2	0.18	0.19	0.19	0.20	0.21	0.22	0.23	0.24	0.26	0.30	0.12
3	0.19	0.20	0.21	0.22	0.24	0.26	0.28	0.30	0.32	0.35	0.16
4	0.22	0.24	0.26	0.28	0.30	0.32	0.34	0.37	0.40	0.46	0.24
5	0.28	0.31	0.33	0.35	0.38	0.41	0.45	0.50	0.56	0.65	0.37
6	0.33	0.36	0.39	0.42	0.46	0.50	0.54	0.61	0.70	0.83	0.50
7	0.45	0.49	0.54	0.58	0.62	0.66	0.71	0.79	0.91	1.18	0.73
8	0.64	0.69	0.75	0.80	0.85	0.90	0.98	1.15	1.37	1.58	0.94
9	1.01	1.09	1.18	1.27	1.36	1.42	1.52	1.76	1.99	2.49	1.48
10 High	3.88	2.93	2.62	2.44	2.41	2.39	2.47	2.71	2.96	3.14	-0.74

Panel E: Portfolio expected excess return spreads, VW (%)

V	H	V 10-1 within each H decile									
10-1	10-1	1	2	3	4	5	6	7	8	9	10
-0.59	1.66	0.07	0.12	0.16	0.22	0.34	0.47	0.68	0.89	1.42	-0.34

Table 12: Variants of the baseline model

This table reports the results for different variants of the baseline model. All results are averages across 1000 simulations.

In the "Regressions" column, the four numbers in each line report the slope coefficients from the cross-sectional regressions of the expected excess return $R_i - R_f$ in percentage on $\log V_i$ only, on $\log H_i$ only, and on both. We use log to denote natural logarithm, and i is the firm subscript.

In the "Portfolio sorting" column, the numbers in each line are equal-weighted expected excess return spreads. Stocks are sorted by size V into deciles, and we report the expected excess return spread between the largest and the smallest V deciles. Stocks are then sorted by idiosyncratic volatility H into deciles, and we report the expected excess return spread between the highest and the lowest H deciles. Finally, stocks are first sorted into H deciles, and then, within each H decile, sorted into V deciles. We report the expected excess return spreads between the largest and the smallest V portfolios within H deciles 1 to 10.

Table 12: continued.

Regressions: $R_i - R_f$(%) regressed on $\log V_i$	Regressions: $R_i - R_f$(%) regressed on $\log H_i$	Regressions: $R_i - R_f$(%) regressed on $\log V_i$ and $\log H_i$		Portfolio sorting: Expected excess return spread, EW (%) V 10-1	H 10-1	V 10-1 within each H decile: 1	2	3	4	5	6	7	8	9	10
0. Baseline															
-0.092	0.922	0.112	1.072	-0.68	2.66	0.07	0.12	0.16	0.24	0.37	0.50	0.73	0.94	1.48	-0.74
1. $\lambda = 1$															
-0.786	3.369	-0.008	3.355	-5.99	8.88	-0.01	-0.01	-0.02	-0.02	-0.02	-0.03	-0.06	-0.11	-0.17	-24.78
2. $\rho = 0$															
0.173	1.789	0.177	1.792	1.09	5.02	0.08	0.13	0.20	0.28	0.39	0.50	0.70	1.16	2.15	5.64
3. Three stocks															
-0.135	1.376	0.166	1.596	-0.99	3.94	0.08	0.15	0.27	0.37	0.48	0.65	0.94	1.28	1.56	-0.88
4. Ten stocks															
-0.038	0.456	0.065	0.543	-0.28	1.37	0.02	0.05	0.08	0.12	0.16	0.22	0.33	0.50	0.84	-0.17
5. No shorting															
-0.079	0.917	0.128	1.088	-0.58	2.62	0.20	0.23	0.26	0.32	0.46	0.59	0.82	1.06	1.58	-0.76
6. Borrowing < 30%															
-0.087	0.920	0.117	1.077	-0.64	2.64	0.10	0.14	0.19	0.27	0.39	0.54	0.76	0.98	1.52	-0.75
7. With mutual funds															
-0.033	0.392	0.055	0.465	-0.24	1.03	0.14	0.17	0.21	0.25	0.29	0.34	0.40	0.53	0.73	-1.14